The Husermet Project – developments to allow large-scale epidemiological and metabolomics studies of the human population

Warwick B Dunn[1], David Broadhurst[2], Paul Begley[1], Eva Zelena[1], Sue Francis-McIntyre[1], Nadine Anderson[1], Marie Brown[1], Andrew Vaughan[1], Wanchang Lin[1], Joshua D Knowles[1], Antony Halsall[1], John N Haselden[3], Andrew W Nicholls[3], Ian D Wilson[4], Douglas B Kell[1], Royston Goodacre[1] & The Human Serum Metabolome (HUSERMET) Consortium[1,2,3,4]

[1] University of Manchester, UK; [2] University of Alberta, Canada; [3] GlaxoSmithKline, UK; [4] AstraZeneca, UK

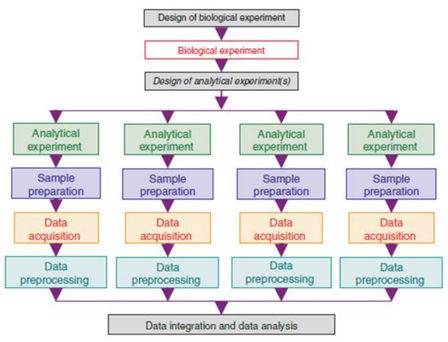
http://husermet.org/index.php

INTRODUCTION

- **The Husermet (HUman SERum METabolome) project was funded to fulfil multiple objectives related to the study of human health, ageing and disease risk factors.**
- **One objective was to characterise the human serum metabolome through a large-scale epidemiological investigation applying untargeted metabolomics.**
- **To fulfil this objective significant development and validation of experimental design and Standard Operating Procedures (SOPs) for sample collection, sample preparation, data acquisition, data pre-processing and quality assurance were required and were successfully performed applying innovative developments. A number of developments are highlighted below in the workflow applied in The Husermet project.**
- **These methods, applying chromatography-mass spectrometry (XC-MS) platforms, have for the first time allowed large-scale epidemiological studies to be performed with XC-MS platforms; metabolomic and clinical biochemistry data has been acquired for greater than 3000 subjects. These data are being integrated with NMR data collected for the same study population.**

EXPERIMENTAL DESIGN IN LARGE-SCALE EPIDEMIOLOGICAL STUDIES

- In large-scale biological studies samples are collected at multiple sites and over many months or years
- To perform metabolomic studies the large biological study is separated in to smaller metabolomic studies followed by data integration after data acquisition – this provides data of the highest quality from XC-MS and NMR platforms
- Appropriate experimental design is required to ensure that the biological study population is represented in each of the metabolomic studies (for example, if the Male:Female ratio is 60:40 in the biological study then this should be observed in each of the metabolomic studies). An optimised algorithm was developed by Joshua Knowles to provide appropriate experimental design.



STANDARD OPERATING PROTOCOLS HAVE BEEN PUBLISHED IN NATURE PROTOCOLS AND ARE APPLIED IN MULTIPLE STUDIES IN ACADEMIA AND INDUSTRY

PROTOCOL

Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry

Warwick B Dunn[1–3], David Broadhurst[2,4], Paul Begley[2], Eva Zelena[2], Sue Francis-McIntyre[2], Nadine Anderson[2], Marie Brown[2], Joshua D Knowles[5], Antony Halsall[2], John N Haselden[6], Andrew W Nicholls[6], Ian D Wilson[7], Douglas B Kell[2], Royston Goodacre[4,2] & The Human Serum Metabolome (HUSERMET) Consortium

[1]Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, UK. [2]School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, UK. [3]Centre for Advanced Discoveries and Experimental Therapeutics, Manchester Biomedical Research Centre and School of Biomedicine, Manchester, UK. [4]Department of Medicine, Kate Group Centre for Pharmacy & Health, University of Alberta, Edmonton, Alberta, Canada. [5]School of Computer Science, The University of Manchester, Manchester, UK. [6]Department of Investigative Preclinical Toxicology, GlaxoSmithKline, Hertfordshire, UK. [7]Department of Clinical Pharmacology, Drug Metabolism and Pharmacokinetics, AstraZeneca, Cheshire, UK. Correspondence should be addressed to W.B.D. (warwick.dunn@manchester.ac.uk).

Published online 30 June 2011; doi:10.1038/nprot.2011.335

Metabolism has an essential role in biological systems. Identification and quantitation of the compounds in the metabolome is defined as metabolic profiling, and it is applied to define metabolic changes related to genetic differences, environmental influences and disease or drug perturbations. Chromatography–mass spectrometry (MS) platforms are frequently used to provide the sensitive and reproducible detection of hundreds to thousands of metabolites in a single biofluid or tissue sample. Here we describe the experimental workflow for long-term and large-scale metabolomic studies involving thousands of human samples with data acquired for multiple analytical batches over many months and years. Protocols for serum- and plasma-based metabolic profiling applying gas chromatography–MS (GC-MS) and ultraperformance liquid chromatography–MS (UPLC-MS) are described. These include biofluid collection, sample preparation, data acquisition, data pre-processing and quality assurance. Methods for quality control–based robust LOESS signal correction to provide signal correction and integration of data from multiple analytical batches are also described.

INTRODUCTION

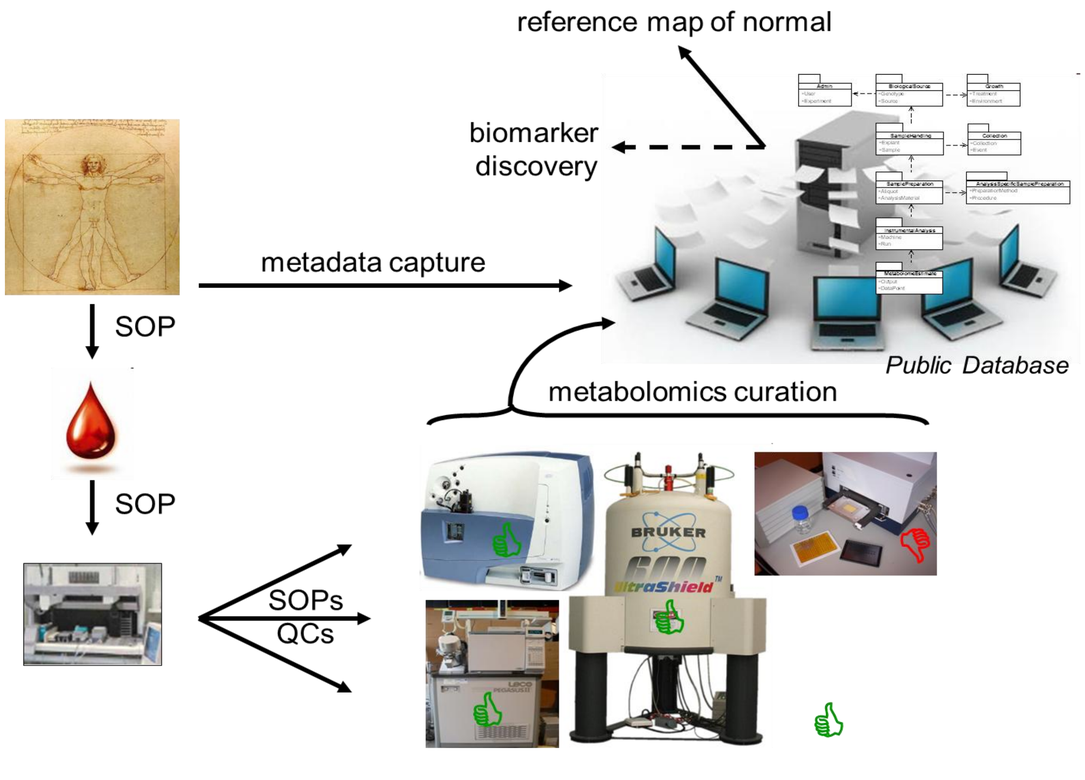
Systems biology is focused on the study of biological components and, more importantly, their complex interactions to define the emergent properties of biological systems[1–3]. Metabolomics[4–6], and the associated field of metabonomics[7], are core areas of systems biology research that are focused on the holistic study of low-molecular-weight organic and inorganic (typically < 1,500 Da) metabolites. Metabolites have an important role in biological systems. They are the building blocks for many other biological components (e.g., proteins, RNA, DNA and cell walls), they are central in intermediary metabolism, they provide many necessities for life (e.g., ATP for energy release) and they have an active role in regulation and signaling. Primary and rapid responses to environmental perturbations are generally, but not exclusively, metabolically focused and are followed by changes at the transcriptional and translational levels.

Metabolomics is applied to the study of microbes[8–10] and plants[11,12], and mammalian[13–16] and environmental systems[17]. In mammals these applications include the study of human diseases to define pathophysiological processes and discover biomarkers[14,18–21], the study of drug toxicity and efficacy[22–24], the study of the interaction of environment and genotype (e.g., nutrigenomics[25,26]) and the study of lipids (lipidomics)[27]. Typically, studies are inductive rather than deductive, and are designed for hypothesis generation or knowledge discovery. This approach starts from a position of limited biological knowledge with the objective to acquire and interrogate data related to a wide and diverse range of metabolites in the metabolome[28]. Holistic studies using MS or nuclear magnetic resonance (NMR) spectroscopy are typically defined as metabolic profiling. Metabolomics offers a number of benefits compared with other '-omic' strategies, with the most advantageous being its close biological proximity to the phenotype of the system and hence the rapid observation of system perturbations in the metabolome. Although analytical platforms are expensive (£100,000s), costs per sample are low (£10s) and, when combined with the high-throughput nature of metabolic profiling (analysis times are typically <30 min), these allow metabolic profiling to be applied to screen large sample sets in a high-throughput approach and at low total cost compared with other '-omic' platforms (e.g., transcriptomics or proteomics).

Metabolic profiling studies have been performed using a range of analytical platforms[29] including gas (GC) or liquid (LC) chromatography (and variants of LC such as UPLC) coupled to MS[30–35], capillary electrophoresis-MS[36,37], NMR spectroscopy[38–40], infrared and Raman spectroscopies[41], electrochemical detectors[42] and direct infusion (or direct injection) MS[43,44]. Of these, chromatography-MS and NMR spectroscopy are the most widely applied and offer different advantages and disadvantages in their application. Owing to the complexity and size of mammalian metabolomes and the diverse physical and chemical properties of metabolites, no single analytical platform can be applied to detect all metabolites in a biological sample[28]. The metabolomics community have realized that the application of multiple analytical platforms in metabolomics is an appropriate strategy to increase the coverage of detected metabolites. For example, the HUSERMET project is applying GC-MS, UPLC-MS and NMR spectroscopy to the epidemiological study of human serum followed by data integration (http://www.husermet.org/).

DEVELOPMENT OF PROTOCOLS FOR SAMPLE COLLECTION AND PREPARATION AND DATA ACQUISITION

- SOPs have been developed for each part of the workflow and include collection of serum and plasma from large populations, sample storage, sample preparation, data acquisition on multiple platforms and data pre-processing
- SOPs for sample collection have allowed greater than 3000 single time-point subject samples to be collected at multiple sites across the UK in a standard manner including sample labelling and identification
- SOPs for sample preparation and analysis have allowed multiple analysts to perform each of the different components of the workflow in a standard manner across many years to allow integration of data from different metabolomic studies.
- SOPs for data pre-processing have allowed us to apply similar processes to multiple different small- and large-scale metabolomic studies



THE APPLICATION OF QUALITY CONTROL SAMPLES

- We have analysed a single representative serum sample intermittently (every 4-7th injection) in all metabolomic studies. This provides a single point of reference across each metabolomic study and for the single biological study
- QC samples are essential in long-term biological experiments to provide:
- 1. equilibration of platforms following routine maintenance
- 2. correction of signal drift within a single metabolomic study
- 3. continuous quality assurance of data for short-term analytical blocks
- 4. integration of multiple analytical blocks
- Quality Control-Robust Loess Signal Correction (QC-RLSC) algorithm for steps 2-4 was designed by Dr David Broadhurst.